# Speaker Verification Using SVM

**Mr. Rastoceanu Florin / Mrs. Lazar Marilena**
Military Equipment and Technologies Research Agency
Aeroportului Street, No. 16, CP 19 OP Bragadiru
077025, Ilfov
ROMANIA

email: rastoceanu_florin@yahoo.com / mnvlazar@yahoo.com

## ABSTRACT

*In this paper, we describe an application of speaker verification using Romanian vowels as speaker's models in case of a small Romanian language database. Afterwards the models are classified with the powerful technique named SVM.*

## 1. INTRODUCTION

If the XX century was the speed century, the one that just begin is the communication century. Because the communication is vital in many areas, the people effort was concentrated in building large communications channels that can transmit more and more information in a shorter time. In the present this objective is almost accomplished and the main effort now is to protect the information that flows through this channels. This is very important, because we know that today the most used communication channels are public, like internet or electromagnetic waves. The first step in protecting this information is the authentication. Biometrics are better methods for authentication and that is the reason that many application use biometric methods. The biometric methods used in present with good results are fingerprint identification, iris scan, face recognition and hand geometry biometrics. Nevertheless those methods needs important resources or are difficult to use. To overstep those limits, person voice could be used for authentication. For example in telephony application the required resources are provided by the phone itself.

Speaker recognition can be used in many areas, like:

- homeland security: airport security, strengthening the national borders, in travel documents, visas;
- enterprise-wide network security infrastructures;
- secure electronic banking;
- investing and other financial transactions;
- retail sales, law enforcement;
- health and social services.

Automatic speaker recognition systems have a wide range of potential applications in Army environments, as well:

- verify the identity of users of various communication channels;
- provide access control to restricted areas, equipment and information;
- verify computer users through terminals accepting voice input;
- counter - terrorism measures. A voice recognition system can be used in identifying an unknown voice recording intercepted by the authorities.

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

| | | |
|---|---|---|
| 1. REPORT DATE<br>**NOV 2010** | 2. REPORT TYPE<br>**N/A** | 3. DATES COVERED<br>**-** |

4. TITLE AND SUBTITLE
**Speaker Verification Using SVM**

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
**Military Equipment and Technologies Research Agency Aeroportului Street, No. 16, CP 19 OP Bragadiru 077025, Ilfov ROMANIA**

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release, distribution unlimited**

13. SUPPLEMENTARY NOTES
**See also ADA564697. Information Assurance and Cyber Defence (Assurance de l’information et cyberdefense). RTO-MP-IST-091**

14. ABSTRACT
**In this paper, we describe an application of speaker verification using Romanian vowels as speakers models in case of a small Romanian language database. Afterwards the models are classified with the powerful technique named SVM.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | **SAR** | **6** | |
| **unclassified** | **unclassified** | **unclassified** | | | |

The paper is organized as follows: After introduction, in the second section it is given a brief introduction to the theory of SVMs. In the third section it is described an experiment using SVMs for a speaker verification application using Romanian vowels. We conclude with the results obtained by application describe above and future work that shall done for increasing the performance of the system.

# 2. SUPPORT VECTOR MACHINES

The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data [1]. The mapping function can be either a classification or a regression function. For classification, nonlinear kernel functions are often used to transform input data to a high-dimensional feature space in which the input data become more separable compared to the original input space. Maximum-margin hyperplanes are then created. The model thus produced depends on only a subset of the training data near the class boundaries.

## 2.1 Linear Case

Consider the problem of separ1ating the set of $N$ training vectors $\{(x^1,y^1), \ldots, (x^n,y^n)\}$, $x \in \Re^m$, belonging to two different classes $y_i \in \{-1, 1\}$. The goal is to find the liniar decision fuction D(x) and the separation plane H.

$$H: < \boldsymbol{w} , \boldsymbol{x} > + \boldsymbol{b} = 0 \qquad (1)$$

$$D(x) = sign\ (\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b}) \qquad (2)$$

where ***b*** is the distance of the hyperplane from the origin and ***w*** is the normal to the decision region.

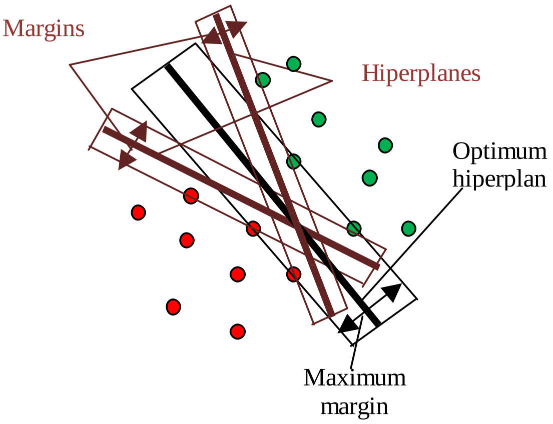

**Figure 1: Separation hyperplanes**

Let the “margin” of the SVM be defined as the distance from the separating hyperplane to the closest two classes. The SVM training paradigm finds the separating hyperplane which gives the maximum margin. The margin is equal to 2/||w||. Once the hyperplane is obtained, all the training examples satisfy the following inequalities [2] :

$$x_i \cdot w + b \geq +1 \qquad for\ y_i = +1 \qquad (3)$$

$$x_i \cdot w + b \geq -1 \qquad for\ y_i = -1 \qquad (4)$$

We can summarize the above procedure to the following:

$$\begin{aligned} &\textit{Minimize} && L(w)=\frac{1}{2}\|w\|^2 \\ &\textit{Subject to} && y_i(x_i\cdot w+b)\geq +1,\quad i=1,2,\ \dots\ ,N \end{aligned} \qquad (5)$$

## 2.2 Non-Linear Case

Real-world classification problems typically involve data that can only be separated using a nonlinear decision surface. Optimization on the input data in this case involves the use of a kernel-based transformation who transform data in a higher dimensional space (feature space) in which data are linear separable.

$$k(x_i,x_j)=\Phi(x_i)\cdot\Phi(x_j) \qquad (6)$$

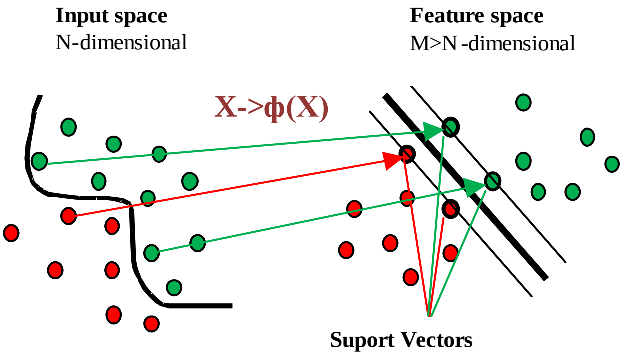

**Figure 2: SVM principle**

Kernels allow a dot product to be computed in a higher dimensional space without explicitly mapping the data into these spaces. The kernels used in our application are:

**Table 1: Kernels used in experiments**

| | |
|---|---|
| RBF | $\exp(g\times(x-x_i)^2)$ |
| Polynomial | $(s\times(x\cdot x_i)+c)^d$ |

# 3. SPEAKER VERIFICATION METHOD

The next figure shows the diagram of the text-independent speaker verification application realized by the authors using the SVM approach.

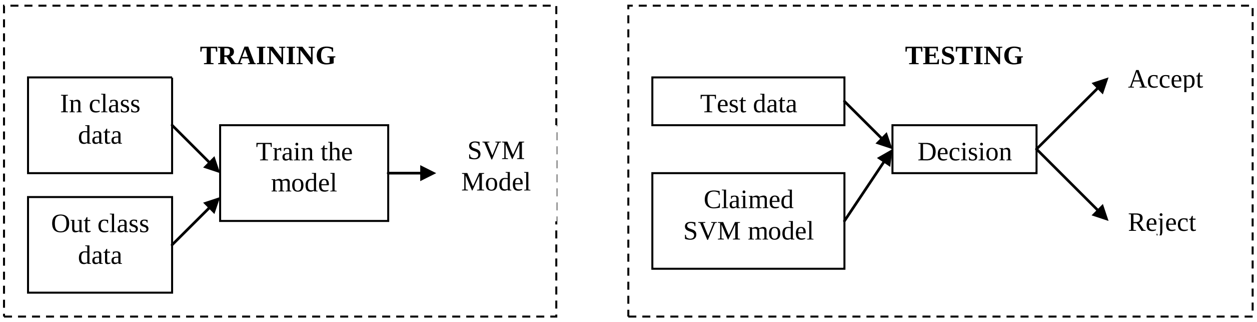

**Figure 3: SVM speaker verification method**

A speaker verification system is composed of two distinct phases, a training phase and a test phase. In the training phase the SVM models corresponding to each speaker are created. For each speaker are created a number of 7 models, one for each Romanian vowel. For training this models are used "in class data" vowels extracted from current speaker and "out class data" vowels extracted from the other speakers. In testing phase the "testing data" are compared with the claimed SVM model and a decision is made. The Equal Error Rate (EER) is used to measure the system performance in all our evaluations. For SVM implementation we use LIBSVM [3], a library for support vector machines classification and regression, developed by National Taiwan University.

## 4. EXPERIMENTS AND RESULTS

The evaluation was carried out on a small database with 10 speakers (2 female and 8 male). A number of 50 different sentences are spoken by each speaker. From this sentences are extracted vowels used for experiment. According with the frequency of its appearance in this sentences are extracted a number of different samples according with vowel's apparition in spoken phrases. The feature extracted from this vowels were 12 LPC coefficients concatenated with 12 delta LPC coefficients (total of 24 coefficients) and a forty dimensional feature vector composed by 12 mel-cepstrum coefficients, log energy, 0th cepstral coefficient, delta and delta-delta coefficients. A set of features corresponding with 80% from the total number of vowels extracted from this sentences were used in the training process and the other 20% were used for testing.

Experiments were carried out to compare the method performances using different types of kernel functions, feature extracted and vowels in a SVM implementation.

In the first stage the comparison are made against the two types of coefficients and kernels used in SVM implementation. For this purpose are used LPC and MFCC coefficient. As a kernel functions are used RBF and Polynomial (degree 3 and 4). The results are presented in figure 4 and table II. We can observe from this that the coefficients with the best results are obtain with MFCC and Polynomial (degree 3) kernel function, but good results are obtain too, using MFCC coefficient with Polynomial (degree 2) and RBF kernel.

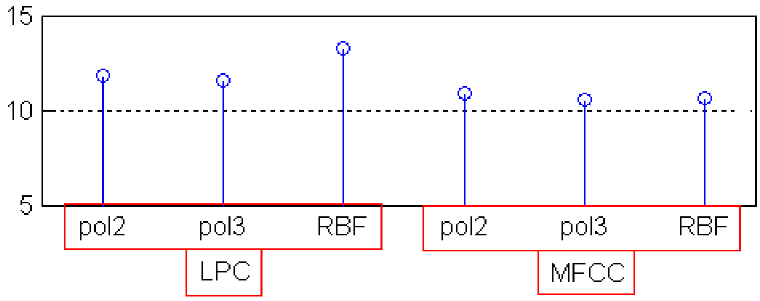

**Figure 4: Mean EER (for all speakers and vowels) for different kernels and coefficients**

**Table 2: Mean EER (for all speakers and vowels) for different kernels and coefficients**

| **Methods** | **Mean EER** |
|---|---|
| LPC+Pol2 | 11.86 |
| LPC+Pol3 | 11.52 |
| LPC+RBF | 13.29 |
| MFCC+Pol2 | 10.88 |
| MFCC+Pol3 | 10.54 |
| MFCC+RBF | 10.67 |

Using the results mentioned above, in the second phase, the comparisons are made using only Polynomial (degree 3) kernel and MFCC coefficients. In this phase the experiments show what are the vowel with the best results (figure 5), and using this vowel, what are the results obtained by each speaker (figure 6).

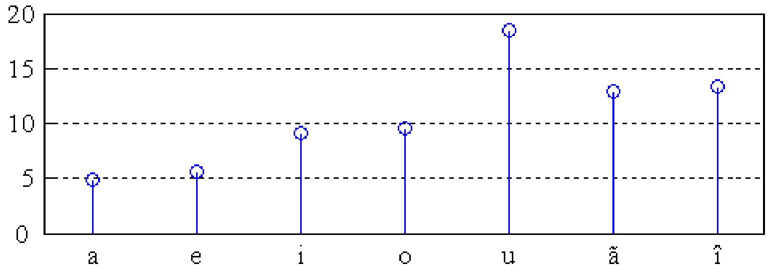

**Figure 5: Mean EER (for all speakers) for Romanian vowels obtained with Polynomial (degree 3) kernel and MFCC coefficients**

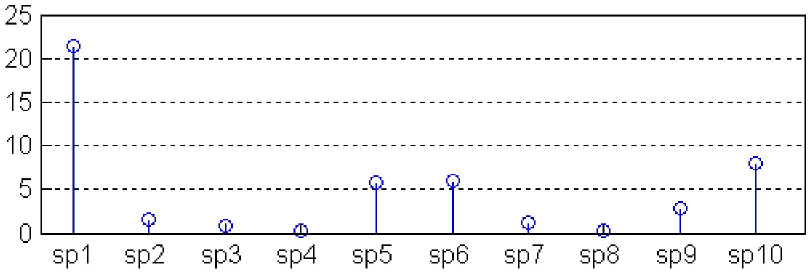

**Figure 6: EER for speakers obtained with Polynomial (degree 3) kernel and MFCC coefficients for vowel "a"**

## 5. CONCLUSIONS

In this paper we describe a method for speaker verification implemented in SVM with SVMLib and a database, recorded in a laboratory, with 10 speakers and 500 sentences. For that purpose, we used LPC and MFCC coefficients as features extracted from Romanian vowels. For SVM we used RBF and polynomial kernels (degree 2 and 3). Our conclusion was that using polynomial (degree 3) kernel, MFCC coefficients and "a" vowel we obtained an EER equal to 4.82, that is a good result.

The differences between the error rates for speakers and vowels are big. For vowels, one supposition is that the database is small and some vowels are better training that the other. Certainly, all these results can be improved using a bigger professional database.

[1] V. Vapnik. "Three remarks on the support vector method of function estimation", In *Advances in Kernel Methods - Support Vector Learning*, MIT Press, 1999

[2] Nello Cristianini and John Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press 2000

[3] www.csie.ntu.edu.tw/~cjlin/libsvm/ - a library for support vector machines classification and regression, developed by National Taiwan University